

# Nonparametric estimation of the conditional tail index and extreme quantiles under random censoring

Pathé NDAO

*LERSTAD, Université Gaston Berger, Saint Louis, Sénégal.*

*Email: ndao.pathe@yahoo.fr*

Aliou DIOP

*LERSTAD, Université Gaston Berger, Saint Louis, Sénégal.*

*Email: aliou.diop@ugb.edu.sn*

Jean-François DUPUY<sup>†</sup>

*IRMAR-Institut National des Sciences Appliquées de Rennes, France.*

*Email: Jean-Francois.Dupuy@insa-rennes.fr*

**Abstract.** In this paper, we investigate the estimation of the tail index and extreme quantiles of a heavy-tailed distribution when some covariate information is available and the data are randomly right-censored. We construct several estimators by combining a moving-window technique (for tackling the covariate information) and the inverse probability-of-censoring weighting method, and we establish their asymptotic normality. A comprehensive simulation study is conducted to evaluate the finite-sample performance of the proposed estimators and to identify their application scope.

**Keywords:** Random censoring, conditional extreme value index, conditional extreme quantiles, heavy-tailed distribution, moving window, simulations.

## 1. Introduction

Let  $Y_1, \dots, Y_n$  be a sequence of independent and identically distributed replicates of a random variable  $Y$ . One question of great practical interest in many domains (reliability, hydrology, insurance, meteorology...) is to estimate extreme quantiles of the distribution of  $Y$  that is, quantities of the form

$$F^{\leftarrow}(1 - \alpha) = \inf\{y : F(y) \geq 1 - \alpha\}$$

where  $\alpha$  is so small that this quantile falls beyond the range of the observed data  $Y_1, \dots, Y_n$ . This problem has been widely investigated so far. It involves the

<sup>†</sup>Corresponding author

estimation of the so-called extreme value index (or tail index)  $\gamma$ . This parameter drives the tail heaviness of the distribution of  $Y$  and thus plays a central role in the analysis of extremes, making its estimation a crucial issue. Detailed accounts on extreme value theory (and in particular on the estimation of the extreme value index and extreme quantiles) can be found, for example, in [2, 9, 14]. Here, we consider the situation where some covariate information  $X$  is available to the investigator, and the distribution of  $Y$  depends on  $X$ . Then for every  $X = x$ , we consider the problem of estimating the conditional extreme value index  $\gamma(x)$  and conditional extreme quantiles  $F^{\leftarrow}(1 - \alpha|x) = \inf\{y : F(y|x) \geq 1 - \alpha\}$  of the distribution  $F(\cdot|x)$  of  $Y$  given  $x$ . Several papers already address the estimation of the conditional extreme value index and conditional extreme quantiles (see for example [10, 11, 13] and the references therein). In the present paper, we consider these issues in the more complicated situation where the observations  $Y_1, \dots, Y_n$  are randomly right-censored. Censoring occurs for example in the statistical analysis of event time data, where  $Y$  represents the time elapsed from some time origin until the occurrence of some event of interest (death of a patient, ruin of a company...) and  $X$  represents some available covariate information (biological markers recorded on a patient, economic characteristics of a company...). If censoring is present, the observations consist of triplets  $(Z_i, \delta_i, X_i)$ ,  $i = 1, \dots, n$ , where  $Z_i = \min(Y_i, C_i)$ ,  $\delta_i = 1_{\{Y_i \leq C_i\}}$ ,  $1_{\{\cdot\}}$  is the indicator function, and  $C_i$  is a random censoring time for the  $i$ -th individual, that provides a lower bound on  $Y_i$  if  $\delta_i = 0$ . The estimation of the extreme value index and extreme quantiles from censored data has been considered, among others, by [3, 5, 8, 12] when there is no covariate information  $X$ . Here, we consider the estimation of these quantities when both censoring and covariate information are present.

We first construct several estimators of the conditional tail index  $\gamma(x)$  of the distribution of  $Y$  given  $x$ , and we establish their asymptotic normality. Our construction combines a moving-window approach (such as developed in [10] in the uncensored case) with the inverse probability-of-censoring weighting (IPCW) principle (such as used in [8] for estimating the unconditional extreme value index with censoring). Then, we construct Weissman-type estimators of the conditional extreme quantile  $F^{\leftarrow}(1 - \alpha|x)$  under censoring. We establish the asymptotic normality of the proposed estimators. Finally, we conduct a simulation study to evaluate the finite-sample performance of all these estimators. The rest of the paper is organized as follows. In Section 2, we set the model and some useful notations. The proposed estimators are constructed in Section 3. In Section 4, we investigate their asymptotic properties. The proofs are postponed to the appendix. The results of a comprehensive simulation study are reported in Section 5. A discussion and some perspectives conclude the paper.

## 2. Model and notations

Let  $(Y_1, \dots, Y_n)$  be  $n$  independent copies of a non-negative random variable  $Y$  and  $(x_1, \dots, x_n)$  be the corresponding values of some  $p$ -dimensional controlled

covariate  $X \in \mathcal{X}$ , where  $\mathcal{X}$  denotes a compact set in  $\mathbb{R}^p$ . We assume that each  $Y_i$  can be right-censored by a non-negative random variable  $C_i$  (the  $C_i$ 's are defined on the same probability space  $(\Omega, \mathcal{C}, \mathbb{P})$  as the  $Y_i$ 's, and are assumed independent of each other), such that we really observe the  $n$  independent triplets  $(Z_i, \delta_i, x_i), i = 1, \dots, n$ , where  $Z_i = \min(Y_i, C_i)$  and  $\delta_i = 1_{\{Y_i \leq C_i\}}$ .  $Y_i$  and  $C_i$  are assumed to be independent. In the sequel,  $d$  will denote the Euclidean distance on  $\mathbb{R}^p$ , and  $\xrightarrow{\mathcal{D}}$  will denote the convergence in distribution.

For every  $x \in \mathcal{X}$ , we denote by  $F(\cdot|x)$  and  $G(\cdot|x)$  respectively the conditional cumulative distribution functions of  $Y$  and  $C$  given  $X = x$ . We assume that for every  $x \in \mathcal{X}$ ,  $F(\cdot|x)$  and  $G(\cdot|x)$  belong to the domain of attraction of Fréchet distributions with shapes  $\gamma_1(x)$  and  $\gamma_2(x)$  respectively. This implies that  $F(\cdot|x)$  and  $G(\cdot|x)$  can be written as

$$F(u|x) = 1 - u^{-1/\gamma_1(x)} L_1(u, x) \quad \text{and} \quad G(u|x) = 1 - u^{-1/\gamma_2(x)} L_2(u, x),$$

where  $\gamma_1(\cdot)$  and  $\gamma_2(\cdot)$  are unknown positive functions of the covariate  $x$  (referred to as conditional tail index functions), and for every  $x \in \mathcal{X}$ ,  $L_1(\cdot, x)$  and  $L_2(\cdot, x)$  are slowly varying at infinity that is, for every  $\lambda > 0$ ,

$$L_i(\lambda u, x)/L_i(u, x) \longrightarrow 1 \quad \text{as} \quad u \longrightarrow \infty, \quad i = 1, 2.$$

Note that by independence of  $Y$  and  $C$ , the conditional cumulative distribution function  $H(\cdot|x)$  of  $Z$  given  $X = x$  is also heavy-tailed, with conditional extreme value index  $\gamma(x) = \gamma_1(x)\gamma_2(x)/(\gamma_1(x) + \gamma_2(x))$ . To see this, note that for every  $u$  and  $x$ ,

$$\begin{aligned} 1 - H(u|x) &= (1 - F(u|x))(1 - G(u|x)) \\ &= u^{-1/\gamma_1(x)} L_1(u, x) u^{-1/\gamma_2(x)} L_2(u, x) \\ &= u^{-1/\gamma(x)} L(u, x) \end{aligned}$$

where  $\gamma(x)$  is as above and  $L(u, x) = L_1(u, x)L_2(u, x)$ . Moreover,

$$\lim_{u \rightarrow \infty} \frac{L(\lambda u, x)}{L(u, x)} = \lim_{u \rightarrow \infty} \frac{L_1(\lambda u, x)}{L_1(u, x)} \frac{L_2(\lambda u, x)}{L_2(u, x)} = 1.$$

Finally, let  $q(\alpha, x)$  be the conditional quantile of order  $1 - \alpha$  ( $\alpha \in (0, 1)$ ) of  $F(\cdot|x)$ , defined by  $F(q(\alpha, x)|x) = 1 - \alpha$ . Given a sample of observations  $(Z_1, \delta_1, x_1), \dots, (Z_n, \delta_n, x_n)$ , our aims are to build and evaluate pointwise estimators of the conditional tail index function  $\gamma_1(x)$  and conditional extreme quantiles  $q(\alpha, x)$ .

### 3. The proposed estimators

Several estimators have been proposed for the extreme value index and extreme quantiles when either some covariate information is available or the  $Y_i$ 's are censored. When both censoring and covariates are present, we propose to estimate

these quantities by combining a moving-window approach (proposed in [10] for estimating the conditional tail index function without censoring) with the IPCW principle (used for example in [8] for estimating the extreme value index in a model without covariate, see also [5]). We first construct a pointwise estimator of the conditional tail index function  $\gamma_1(x)$ .

### 3.1. Estimation of the conditional tail index function

If  $x \in \mathcal{X}$  and  $r > 0$ , let  $B(x, r)$  denote the ball in  $\mathbb{R}^p$  with center  $x$  and radius  $r$  that is,  $B(x, r) = \{t \in \mathbb{R}^p, d(x, t) \leq r\}$ . Let  $h_{n,x}$  be a positive sequence tending to 0 as  $n$  tends to infinity, and define  $m_{n,x} := \sum_{i=1}^n 1_{\{x_i \in B(x, h_{n,x})\}}$  (respectively  $\phi(h_{n,x}) := n^{-1}m_{n,x}$ ) as the number (respectively the proportion) of the observations  $(Z_i, x_i)$  lying in  $[0, \infty) \times B(x, h_{n,x})$ . Let  $Z_{(1)}^x \leq \dots \leq Z_{(m_{n,x})}^x$  be the ordered values of  $Z$  for these observations, and let  $\delta_{(1)}^x, \dots, \delta_{(m_{n,x})}^x$  be the corresponding  $\delta$ 's (that is,  $\delta_{(i)}^x = \delta_j$  if  $Z_{(i)}^x = Z_j$ ). Note that since  $X$  is controlled,  $m_{n,x}$  and  $\phi(h_{n,x})$  are nonrandom numbers. If the  $Z_{(i)}^x$  were not censored, Gardes and Girard (see [10]) propose to estimate  $\gamma_1(x)$  by a Hill-type estimator of the form

$$\hat{\gamma}_{k_x, m_{n,x}}^{(H)}(x) := M_{k_x, m_{n,x}}^{(1)} := \frac{1}{k_x} \sum_{i=1}^{k_x} i \log \left( \frac{Z_{(m_{n,x}-i+1)}^x}{Z_{(m_{n,x}-i)}^x} \right) \quad (3.1)$$

where  $k_x$  is a sequence of integers such that  $1 \leq k_x \leq m_{n,x}$ . Several other estimators of the extreme value index can be adapted to the situation where  $\gamma_1$  is a function of a covariate. For example, one may consider the following adapted version of the moment estimator:

$$\hat{\gamma}_{k_x, m_{n,x}}^{(M)}(x) := M_{k_x, m_{n,x}}^{(1)} + 1 - \frac{1}{2} \left( 1 - \frac{(M_{k_x, m_{n,x}}^{(1)})^2}{M_{k_x, m_{n,x}}^{(2)}} \right)^{-1} \quad (3.2)$$

where

$$M_{k_x, m_{n,x}}^{(2)} := \frac{1}{k_x} \sum_{i=1}^{k_x} \left( i \log \left( \frac{Z_{(m_{n,x}-i+1)}^x}{Z_{(m_{n,x}-i)}^x} \right) \right)^2,$$

or the following adapted version of the UH estimator:

$$\hat{\gamma}_{k_x, m_{n,x}}^{(UH)}(x) := \frac{1}{k_x} \sum_{i=1}^{k_x} \log \left( \frac{Z_{(m_{n,x}-i)}^x \hat{\gamma}_{i, m_{n,x}}^{(H)}(x)}{Z_{(m_{n,x}-k_x)}^x \hat{\gamma}_{k_x, m_{n,x}}^{(H)}(x)} \right). \quad (3.3)$$

The estimators (3.2) and (3.3) extend the estimators proposed in [7] and [4] respectively when there is no covariate information. In what follows, we adapt the estimators (3.1), (3.2), and (3.3) to the case where censoring occurs (note

that these estimators are not consistent for  $\gamma_1(x)$  if they are directly applied to the sample  $(Z_i, \delta_i, x_i), i = 1, \dots, n$ . Indeed, they will converge to the extreme value index  $\gamma(x)$  of the conditional distribution of  $Z$ .

To accomodate censoring, we divide the estimators (3.1), (3.2), and (3.3) by the proportion

$$\hat{p}_x = \frac{1}{k_x} \sum_{i=1}^{k_x} \delta_{(m_{n,x}-i+1)}^x$$

of uncensored observations among the  $k_x$  largest  $Z$ 's in a neighbourhood of  $x$  (a similar idea was used, for example, in [5] and [8] for estimating the extreme value index without covariate). For all  $x \in \mathcal{X}$ , our proposal is thus to estimate  $\gamma_1(x)$  by

$$\hat{\gamma}_{k_x, m_{n,x}}^{(c, \cdot)}(x) := \frac{\hat{\gamma}_{k_x, m_{n,x}}^{(\cdot)}(x)}{\hat{p}_x} \quad (3.4)$$

where the  $\cdot$  in  $\hat{\gamma}_{k_x, m_{n,x}}^{(c, \cdot)}(x)$  and  $\hat{\gamma}_{k_x, m_{n,x}}^{(\cdot)}(x)$  stands for any of the Hill, moment, and UH estimator.

### 3.2. Estimation of conditional extreme quantiles

In this section, we further address the estimation of conditional extreme quantiles  $q(\alpha_{m_{n,x}}, x) \in \mathbb{R}$  of order  $1 - \alpha_{m_{n,x}}$  of the distribution of  $Y$  given  $X = x$ . Such quantiles verify  $1 - F(q(\alpha_{m_{n,x}}, x)|x) = \alpha_{m_{n,x}}$  where  $\alpha_{m_{n,x}} \rightarrow 0$  as  $m_{n,x} \rightarrow +\infty$ . For every  $x \in \mathcal{X}$ , we first consider the following conditional Kaplan-Meier-type estimator, based on the moving-window approach described in the section 3.1:

$$1 - \hat{F}_{m_{n,x}}(y|x) = \prod_{i=1}^{m_{n,x}} \left( \frac{m_{n,x} - i}{m_{n,x} - i + 1} \right)^{\delta_{(i)}^x 1_{\{Z_{(i)}^x \leq y\}}}.$$

Based on this, we propose to estimate  $q(\alpha_{m_{n,x}}, x)$  by the following Weissman-type estimator ([15]):

$$\hat{q}^{(c, \cdot)}(\alpha_{m_{n,x}}, x) = Z_{(m_{n,x}-k_x)}^x \left( \frac{1 - \hat{F}_{m_{n,x}}(Z_{(m_{n,x}-k_x)}^x|x)}{\alpha_{m_{n,x}}} \right)^{\hat{\gamma}_{k_x, m_{n,x}}^{(c, \cdot)}(x)} \quad (3.5)$$

where  $\hat{\gamma}_{k_x, m_{n,x}}^{(c, \cdot)}(x)$  is any of the estimators (3.4). Note that (3.5) extends the conditional extreme quantile estimator proposed in [13] in the situation where there is no censoring.

In the next section, we establish the asymptotic properties of our estimators (3.4) and (3.5).

#### 4. Asymptotic results

We first state some regularity conditions that will be needed for proving our asymptotic results (these conditions are used in [10] to prove the asymptotic normality of the conditional tail index estimator without censoring, see also [5] for the case with censoring but no covariate information). We assume that:

**C1** for every  $x \in \mathcal{X}$ , the conditional distribution functions  $F(\cdot|x)$  and  $G(\cdot|x)$  are absolutely continuous,

**C2** for every  $x \in \mathcal{X}$ , there exists a function  $\rho(x) < 0$  and a regularly varying function  $b(\cdot, x)$  with index  $\rho(x)$  such that for any  $u > 0$ ,

$$\lim_{t \rightarrow \infty} \frac{H^\leftarrow(1 - \frac{1}{tu} | x) / H^\leftarrow(1 - \frac{1}{t} | x) - u^{\gamma(x)}}{b(t, x)} = u^{\gamma(x)} \frac{u^{\rho(x)} - 1}{\rho(x)},$$

The following assumptions are also required (they are similar to the conditions given in [5] for estimating the unconditional extreme value index with censoring). For any  $x \in \mathcal{X}$ , let  $p_x = \gamma_2(x)/(\gamma_1(x) + \gamma_2(x))$ . Assume that as  $n \rightarrow \infty$ ,  $k_x \rightarrow \infty$ ,  $\frac{k_x}{m_{n,x}} \rightarrow 0$ , and:

**C3**  $\sqrt{k_x} b\left(\frac{m_{n,x}}{k_x}, x\right) \rightarrow \lambda(x) < \infty$ ,

**C4**  $\frac{1}{\sqrt{k_x}} \sum_{i=1}^{k_x} \left[ p_x \left( H^\leftarrow\left(1 - \frac{i}{m_{n,x}} | x\right) \right) - p_x \right] \rightarrow \epsilon(x) < \infty$ ,

**C5** letting  $\mathcal{A}(s, t) := \{1 - \frac{k_x}{m_{n,x}} \leq t < 1, |t - s| \leq C \frac{\sqrt{k_x}}{m_{n,x}}, s < 1\}$ , we assume  $\sqrt{k_x} \sup_{\mathcal{A}(s,t)} |p_x(H^\leftarrow(t|x)) - p_x(H^\leftarrow(s|x))| \rightarrow 0$ , for all  $C > 0$ .

We are now in position to state our first main result. Its proof is given in the appendix.

**THEOREM 4.1.** *Let  $x \in \mathcal{X}$ . Assume that the conditions C1-C5 hold and that there exist some functions  $m(\cdot)$  and  $\sigma(\cdot)$  such that  $\sqrt{k_x} \left( \hat{\gamma}_{k_x, m_{n,x}}^{(\cdot)}(x) - \gamma(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(m(x)\lambda(x), \sigma^2(x))$ . Then the following holds:*

$$\begin{aligned} & \sqrt{k_x} \left( \hat{\gamma}_{k_x, m_{n,x}}^{(c, \cdot)}(x) - \gamma_1(x) \right) \\ & \xrightarrow{\mathcal{D}} \mathcal{N} \left( \frac{1}{p_x} (\lambda(x)m(x) - \gamma_1(x)\epsilon(x)), \frac{\sigma^2(x) + \gamma_1(x)^2 p_x (1 - p_x)}{p_x^2} \right). \end{aligned}$$

Considering successively each of the Hill, moment, and UH estimator, we obtain the following corollary:

COROLLARY 4.2. *Under the assumptions of Theorem 4.1, the following holds for every  $x \in \mathcal{X}$ :*

$$\begin{aligned}
& \sqrt{k_x} \left( \hat{\gamma}_{k_x, m_{n,x}}^{(c,H)}(x) - \gamma_1(x) \right) \\
& \xrightarrow{\mathcal{D}} \mathcal{N} \left( \frac{-\gamma_1(x)\epsilon(x)}{p_x} + \frac{\lambda(x)}{p_x(1-\rho(x))}, \frac{\gamma_1^3(x)}{\gamma_1(x)} \right), \\
& \sqrt{k_x} \left( \hat{\gamma}_{k_x, m_{n,x}}^{(c,UH)}(x) - \gamma_1(x) \right) \\
& \xrightarrow{\mathcal{D}} \mathcal{N} \left( \frac{-\gamma_1(x)\epsilon(x)}{p_x} + \frac{\lambda(x)}{p_x(1-\rho(x))}, \frac{\gamma_1^2(x)}{\gamma_1^2(x)}(1 + \gamma_1(x)\gamma(x)) \right), \\
& \sqrt{k_x} \left( \hat{\gamma}_{k_x, m_{n,x}}^{(c,M)}(x) - \gamma_1(x) \right) \\
& \xrightarrow{\mathcal{D}} \mathcal{N} \left( \frac{-\gamma_1(x)\epsilon(x)}{p_x} + \frac{\lambda(x)}{p_x(1-\rho(x))}, \frac{\gamma_1^2(x)}{\gamma_1^2(x)}(1 + \gamma_1(x)\gamma(x)) \right).
\end{aligned}$$

This corollary is easily proved by noting that  $m(x) = (1 - \rho(x))^{-1}$  (for all  $\gamma_{k_x, m_{n,x}}^{(H)}(x)$ ,  $\gamma_{k_x, m_{n,x}}^M(x)$ , and  $\gamma_{k_x, m_{n,x}}^{(UH)}(x)$ ), and that (see [1] and [7]):

$$\sigma^2(x) = \begin{cases} \gamma^2(x) & \text{for } \gamma_{k_x, m_{n,x}}^{(H)}(x) \\ 1 + \gamma^2(x) & \text{for } \gamma_{k_x, m_{n,x}}^{(M)}(x) \\ 1 + \gamma^2(x) & \text{for } \gamma_{k_x, m_{n,x}}^{(UH)}(x). \end{cases}$$

We now turn to the asymptotic properties of the estimator (3.5) of the conditional extreme quantiles. The following additional notations and regularity condition are needed (see [13]):

**C6** for every  $x \in \mathcal{X}$ , the conditional quantile function  $\alpha \in (0, 1) \mapsto q(\alpha, x) \in (0, +\infty)$  is differentiable and the function  $\alpha \in (0, 1) \mapsto \Delta(\alpha, x) = \gamma_1(x) + \alpha(\partial \log q(\alpha, x)/\partial \alpha)$  is continuous and tends to 0 as  $\alpha$  tends to 0.

Let  $\bar{\Delta}(a, x) = \sup_{\alpha \in (0, a)} |\Delta(\alpha, x)|$  and for any  $a \in (0, 1/2)$ , let

$$\omega_n(a) = \sup \left\{ \left| \log \frac{q(\alpha, t)}{q(\alpha, t')} \right| ; \alpha \in (a, 1-a), (t, t') \in (B(x, h_{n,x}))^2 \right\}.$$

We now state our second main result, which gives the asymptotic properties of the estimator (3.5) (its proof is given in the appendix).

THEOREM 4.3. *Assume that the conditions C1-C6 hold. Let  $(\beta_{m_{n,x}})_{n \geq 1} := (1 - \hat{F}_{m_{n,x}}(Z_{(m_{n,x} - k_x)}^x | x))_{n \geq 1}$  and  $(\alpha_{m_{n,x}})_{n \geq 1}$  be a sequence such that  $\alpha_{m_{n,x}} < \beta_{m_{n,x}}$ . Let also  $\xi_{m_{n,x}} = (m_{n,x} \beta_{m_{n,x}})^{1/2} \log(\beta_{m_{n,x}} / \alpha_{m_{n,x}})$ . Assume that as  $n \rightarrow \infty$ , there exists  $\delta > 0$  such that  $(m_{n,x} \beta_{m_{n,x}})^2 \omega_n(m_{n,x}^{-(1+\delta)}) \rightarrow 0$ , and*

$k_x^{1/2} \max\{\xi_{m_{n,x}}^{-1}, \bar{\Delta}(\beta_{m_{n,x}}, x)\} \rightarrow 0$ . Then

$$\frac{\sqrt{k_x}}{\log(\beta_{m_{n,x}}/\alpha_{m_{n,x}})} \log \left( \frac{\hat{q}^{(c,\cdot)}(\alpha_{m_{n,x}}, x)}{q(\alpha_{m_{n,x}}, x)} \right) \\ \xrightarrow{\mathcal{D}} \mathcal{N} \left( \frac{1}{p_x} (\lambda(x)m(x) - \gamma_1(x)\epsilon(x)), \frac{\sigma^2(x) + \gamma_1(x)^2 p_x (1 - p_x)}{p_x^2} \right).$$

From this, one can easily derive the asymptotic distribution of  $\hat{q}^{(c,\cdot)}(\alpha_{m_{n,x}}, x)$  for each particular case (Hill, moment, UH). This proceeds along the same lines as the Corollary 4.2, and is omitted for conciseness.

## 5. Simulation study

In this section, we conduct a comprehensive simulation study to evaluate the performance of the proposed estimators (3.4) and (3.5) of the conditional extreme value index and conditional quantiles. We investigate both the accuracy of these estimators and the quality of the Gaussian approximation of their asymptotic distributions. We identify the application scope of each of these estimators (in terms of the sample size and censoring proportion).

### 5.1. The study design

The simulation design (inspired by [13]) is as follows. We simulate  $R = 1000$  samples of size  $n$  ( $n = 500, 1000, 1500, 2000$ ) of independent replicates  $(Z_i, \delta_i, x_i)$ , where  $Z_i = \min(Y_i, C_i)$  and  $x_i \in [0, 1]$ . The conditional distribution of  $Y_i$  given  $X = x_i$  is Pareto with parameter

$$\gamma_1(x) = .5 (.1 + \sin(\pi x)) \left( 1.1 - .5 \exp \left( -64 (x - .5)^2 \right) \right)$$

and the distribution of  $C_i$  is Pareto with a parameter  $\gamma_2$  chosen to yield the desired censoring percentage  $c$  ( $c$  is successively chosen equal to 10%, 25%, 40%). The pattern of  $\gamma_1(\cdot)$  is given on Figure 1.

FIGURE 1 HERE

For each of the  $R$  samples, we estimate  $\gamma_1(\cdot)$  at  $x = 0.5$  ( $\gamma_1(0.5) = 0.35$ ) by each of the Hill, moment, and UH estimators (3.4). The moving window approach described in Section 3 is used with the ball  $B(0.5, 0.1)$ . Choosing the most appropriate value for  $k_x$  is a difficult issue, and we refer the reader to [6] and [13] for a detailed discussion. A more detailed investigation of this issue in our setting falls out of the scope of the present paper. For illustrative purpose, let  $\hat{\gamma}_{i,m_{n,x}}^{(c,\cdot),j}(x)$  denote the estimate of  $\gamma_1(x)$  obtained in the  $j$ -th sample ( $j = 1, \dots, R$ ) with  $k_x = i$  ( $i = 1, \dots, m_{n,x}$ ). Then we retain the following value for



$$k_x: k_x^{opt} := \underset{1 \leq i \leq m_{n,x}}{\operatorname{argmin}} MSE(\hat{\gamma}_{i,m_{n,x}}^{(c,\cdot)}(x)) = \underset{1 \leq i \leq m_{n,x}}{\operatorname{argmin}} R^{-1} \sum_{j=1}^R (\hat{\gamma}_{i,m_{n,x}}^{(c,\cdot),j}(x) - \gamma_1(x))^2$$

(where MSE stands for mean square error), keeping in mind that this method should be modified in practice since  $\gamma_1(x)$  is unknown. Using this value of  $k_x$ , we calculate, for each of the Hill, moment, and UH estimators and each censoring percentage, the averaged estimates of  $\gamma_1(x)$ , along with their empirical root mean square and mean absolute errors. Finally, for each configuration of the simulation design parameters, we compute confidence intervals of asymptotic level 95% for  $\gamma_1(x)$ , and we obtain the empirical coverage probabilities over the  $R$  intervals (plug-in estimates are obtained for the asymptotic variance). The results are given in Table 1.

TABLE 1 HERE

In order to evaluate the quality of the Gaussian approximation of the asymptotic distribution of the proposed estimator (3.4) of the conditional tail index (at  $x = 0.5$ ), we plot the histograms of the  $R$  Hill, moment, and UH estimates. The histograms are given in the Figures 2 (for  $n = 500$ ) and 3 (for  $n = 1500$ ). For each of Hill, moment and UH, we also represent the averaged estimate of  $\gamma_1(x)$  and the corresponding empirical MSE, as functions of  $k_x$  (see Figure 4). The plots are given for  $c = 10\%, 25\%, 40\%$  and  $n = 1000$  (the graphs for the other values of  $n$  yield similar observations and are therefore omitted).

FIGURES 2, 3, and 4 HERE

Next, we turn to the estimation of the extreme quantile  $q(1/5000, 0.5)$  of order  $1 - 1/5000$  of the conditional distribution of  $Y$  given  $x = 0.5$  ( $q(1/5000, 0.5) \approx 19.70786$ ). For each configuration of the simulation design parameters, we calculate the conditional estimate (3.5), based on the Hill, moment, and UH estimators of the conditional tail index. Then, for each sample size  $n$  and censoring percentage  $c$ , we obtain the averaged value of the  $\hat{q}^{(c,\cdot),j}(1/5000, 0.5)$  ( $j = 1, \dots, R$ ), along with their empirical root mean square and mean absolute errors. Finally, we compute confidence intervals (with asymptotic level 95%) for  $q(1/5000, 0.5)$  and we obtain the empirical coverage probabilities over the  $R$  resulting intervals. Table 2 reports the results.

TABLE 2 HERE

Similarly as above, we evaluate the quality of the Gaussian approximation of the asymptotic distribution of the estimator (3.5) of the extreme quantile  $q(1/5000, 0.5)$ . We plot the histograms of the  $R = 1000$  estimates of  $q(1/5000, 0.5)$  (based on the Hill, moment, and UH estimates of the conditional tail index). The histograms are given in the Figures 5 ( $n = 500$ ) and 6 ( $n = 1500$ ). We also represent the averaged estimate of  $q(1/5000, 0.5)$  and the corresponding empirical MSE as functions of  $k_x$  (see Figure 7), when the conditional tail index is estimated by the Hill, moment, and UH estimators respectively. The plots are given for  $c = 10\%, 25\%, 40\%$  and  $n = 1000$  (the graphs for  $n = 500, 1500, 2000$  yield similar observations and are therefore omitted).

FIGURES 5, 6, and 7 HERE

## 5.2. Results

From the Table 1, the quality of the various considered estimators of  $\gamma_1(x)$  degrades as the censoring percentage increases and the sample size decreases. Note that the empirical coverage probabilities increase as the censoring proportion increases, which comes from the fact that the variance estimate increases (this in turn implies that the confidence intervals become wider, and not more precise). The Hill estimator performs much better than the moment and UH for every configuration of the simulation parameters. In particular, the Hill estimator is much less biased than the two others (this is particularly noticeable when the sample size is moderate) and is more robust to censoring. The Figures 2 and 3 reveal that the Gaussian approximation of the asymptotic distribution of the Hill estimator is reasonably satisfied even for a moderate sample size ( $n = 500$ ). When  $n = 500$  and the censoring percentage is moderate (25%) to large (40%), the distributions of the moment and UH estimators are slightly skewed. The superiority of the Hill estimator is also noticeable on the Figure 4.

From the Table 2, the moment and UH estimators of the conditional extreme quantile appear to be biased, even when the sample size is large, and much less robust to censoring than the Hill-based estimator. From this table also, the Hill estimator provides a satisfactory approximation of the true extreme quantile, even when the censoring is heavy and the sample size is small. Moreover, the moment and UH estimators of the conditional extreme quantile suffer numerical instability, as can be noticed from the upper bounds of the confidence intervals, which can be meaningless when the sample size is moderate ( $n = 500$ ) or the censoring fraction is moderate to large. From the Figures 5 and 6, the moment and UH estimators are strongly skewed in almost all configurations of the simulation design. They appear to be moderately skewed when the sample size is large and the censoring proportion is small (10%). The Hill estimator is moderately skewed when the sample size is small. Its distribution is close to the Gaussian when the sample size is large. Finally, from the Figure 7, we observe that the quantile estimate is rather sensitive to the choice of  $k_x$  unless the censoring fraction is small. From this figure, we also observe the superiority of the Hill estimator over the moment and UH in terms of MSE.

## 6. Discussion and perspectives

In this paper, we have considered the estimation of the tail index and extreme quantiles of a heavy-tailed distribution when some covariate information is available and the data are randomly right-censored. We have constructed several estimators for these quantities, by combining a moving-window approach and the inverse probability-of-censoring weighting method. We have established the asymptotic normality of these estimators. A comprehensive simulation study

was conducted to evaluate and compare their finite-sample performance. The Hill estimators of the conditional tail index and extreme quantiles appear to outperform the moment and UH estimators: the Hill estimators are less biased, more robust to censoring and more stable with respect to the choice of  $k_x$ .

Several issues still deserve attention. In particular, the proposed estimators rely on the smoothing parameter  $h_{n,x}$  and the number  $k_x$  of upper order statistics. A detailed investigation of how one may choose these values in practice is needed, and is the topic for future research. In this work, we considered the case where the covariate  $X$  is controlled. The case where  $X$  is random is the topic for our current investigations.

## Appendix: proofs of theorems

**Proof of Theorem 4.1.** The proof proceeds along the same lines as the proof of Theorem 1 in [8], thus we mention the main steps only. We consider first the following decomposition (for any of the Hill, moment, and UH estimator):

$$\begin{aligned} \sqrt{k_x} \left( \hat{\gamma}_{k_x, m_{n,x}}^{(c, \cdot)}(x) - \gamma_1(x) \right) &= \frac{1}{\hat{p}_x} \sqrt{k_x} \left( \hat{\gamma}_{k_x, m_{n,x}}^{(\cdot)}(x) - \gamma(x) \right) \\ &\quad + \frac{1}{\hat{p}_x} \sqrt{k_x} (\gamma(x) - \gamma_1(x) \hat{p}_x) \\ &= \frac{1}{\hat{p}_x} \sqrt{k_x} \left( \hat{\gamma}_{k_x, m_{n,x}}^{(\cdot)}(x) - \gamma(x) \right) \\ &\quad - \frac{\gamma_1(x)}{\hat{p}_x} \sqrt{k_x} \left( \hat{p}_x - \frac{\gamma_2(x)}{\gamma_1(x) + \gamma_2(x)} \right). \end{aligned} \quad (6.6)$$

Consider first the first term in the right-hand side of (6.6). Under the conditions stated in Section 4, it follows from [10] that as  $n \rightarrow \infty$ ,

$$\sqrt{k_x} \left( \hat{\gamma}_{k_x, m_{n,x}}^{(\cdot)}(x) - \gamma(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(m(x)\lambda(x), \sigma^2(x)).$$

Moreover, for every  $x \in \mathcal{X}$ ,  $\sqrt{k_x}(\hat{p}_x - p_x) \xrightarrow{\mathcal{D}} \mathcal{N}(\epsilon(x), p_x(1 - p_x))$  (the proof is similar to the proof of Theorem 1 in [8] and is therefore omitted) and thus

$$\frac{\sqrt{k_x}}{\hat{p}_x} \left( \hat{\gamma}_{k_x, m_{n,x}}^{(\cdot)}(x) - \gamma(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{m(x)\lambda(x)}{p_x}, \frac{\sigma^2(x)}{p_x^2}\right).$$

Consider now the second term in the right-hand side of (6.6). As mentioned above,  $\sqrt{k_x}(\hat{p}_x - p_x) \xrightarrow{\mathcal{D}} \mathcal{N}(\epsilon(x), p_x(1 - p_x))$ , and thus by independence,

$$\begin{aligned} \sqrt{k_x} \left( \hat{\gamma}_{k_x, m_{n,x}}^{(c, \cdot)}(x) - \gamma_1(x) \right) \\ \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{1}{p_x}(\lambda(x)m(x) - \gamma_1(x)\epsilon(x)), \frac{\sigma^2(x) + \gamma_1(x)^2 p_x(1 - p_x)}{p_x^2}\right). \end{aligned}$$

**Proof of theorem 4.3.** The proof follows the same steps as the proof of Theorem 4.3.3 in [13], hence we only outline the main steps. Letting  $\beta_{m_{n,x}} := 1 - \widehat{F}_{m_{n,x}}(Z_{(m_{n,x}-k_x)}^x|x)$ , we get that

$$\beta_{m_{n,x}} = \prod_{i=1}^{m_{n,x}-k_x} \left( \frac{m_{n,x}-i}{m_{n,x}-i+1} \right)^{\delta_{(i)}^x} \leq 1.$$

Moreover,

$$\frac{k_x}{m_{n,x}} = \prod_{i=1}^{m_{n,x}-k_x} \left( \frac{m_{n,x}-i}{m_{n,x}-i+1} \right) \leq \beta_{m_{n,x}}.$$

It follows that  $k_x \leq m_{n,x}\beta_{m_{n,x}} \leq m_{n,x}$  and thus,  $m_{n,x}\beta_{m_{n,x}} \rightarrow \infty$  as  $n \rightarrow \infty$ . The Theorem 4.3.1 of [13] therefore applies, and together with the delta-method, it implies that

$$(m_{n,x}\beta_{m_{n,x}})^{1/2} \log \left( \frac{Z_{(m_{n,x}-k_x)}^x}{q(\beta_{m_{n,x}}, x)} \right) = O_p(1). \quad (6.7)$$

Now, it follows from (3.5) that

$$\log \widehat{q}^{(c,\cdot)}(\alpha_{m_{n,x}}, x) = \log Z_{(m_{n,x}-k_x)}^x + \widehat{\gamma}_{k_x, m_{n,x}}^{(c,\cdot)}(x) \log \left( \frac{\beta_{m_{n,x}}}{\alpha_{m_{n,x}}} \right)$$

and thus

$$\begin{aligned} & \frac{\sqrt{k_x}}{\log(\beta_{m_{n,x}}/\alpha_{m_{n,x}})} \log \left( \frac{\widehat{q}^{(c,\cdot)}(\alpha_{m_{n,x}}, x)}{q(\alpha_{m_{n,x}}, x)} \right) \\ &= \frac{\sqrt{k_x}}{\log(\beta_{m_{n,x}}/\alpha_{m_{n,x}})} \log \left( \frac{Z_{(m_{n,x}-k_x)}^x}{q(\beta_{m_{n,x}}, x)} \right) \\ & \quad + \sqrt{k_x} \left( \widehat{\gamma}_{k_x, m_{n,x}}^{(c,\cdot)}(x) - \gamma_1(x) \right) \\ & \quad - \frac{\sqrt{k_x}}{\log(\beta_{m_{n,x}}/\alpha_{m_{n,x}})} \left( \log \left( \frac{q(\alpha_{m_{n,x}}, x)}{q(\beta_{m_{n,x}}, x)} \right) + \gamma_1(x) \log \left( \frac{\beta_{m_{n,x}}}{\alpha_{m_{n,x}}} \right) \right) \\ &:= \xi_{1, m_{n,x}} + \xi_{2, m_{n,x}} - \xi_{3, m_{n,x}} \end{aligned}$$

where

$$\begin{aligned}\xi_{1,m_{n,x}} &= \frac{\sqrt{k_x}}{\log(\beta_{m_{n,x}}/\alpha_{m_{n,x}})} \log \left( \frac{Z_{(m_{n,x}-k_x)}^x}{q(\beta_{m_{n,x}}, x)} \right) \\ \xi_{2,m_{n,x}} &= \sqrt{k_x} \left( \hat{\gamma}_{k_x, m_{n,x}}^{(c, \cdot)}(x) - \gamma_1(x) \right) \\ \xi_{3,m_{n,x}} &= \frac{\sqrt{k_x}}{\log(\beta_{m_{n,x}}/\alpha_{m_{n,x}})} \left( \log \left( \frac{q(\alpha_{m_{n,x}}, x)}{q(\beta_{m_{n,x}}, x)} \right) + \gamma_1(x) \log \left( \frac{\beta_{m_{n,x}}}{\alpha_{m_{n,x}}} \right) \right).\end{aligned}$$

Note first that with the notations of Theorem 4.3,

$$\xi_{1,m_{n,x}} = \sqrt{k_x} \xi_{m_{n,x}}^{-1} (m_{n,x} \beta_{m_{n,x}})^{1/2} \log \left( \frac{Z_{(m_{n,x}-k_x)}^x}{q(\beta_{m_{n,x}}, x)} \right).$$

It follows from (6.7) and the assumptions of Theorem 4.3 that  $\xi_{1,m_{n,x}} \rightarrow 0$  in probability as  $n \rightarrow \infty$ . Next, from the Theorem (4.1),  $\xi_{2,m_{n,x}}$  converges in distribution to  $\mathcal{N} \left( \frac{1}{p_x} (\lambda(x)m(x) - \gamma_1(x)\epsilon(x)), \frac{\sigma^2(x) + \gamma_1^2(x)p_x(1-p_x)}{p_x^2} \right)$ . Finally, some calculations yield

$$\xi_{3,m_{n,x}} = -\frac{\sqrt{k_x}}{\log(\beta_{m_{n,x}}/\alpha_{m_{n,x}})} \int_{\alpha_{m_{n,x}}}^{\beta_{m_{n,x}}} \frac{\Delta(u, x)}{u} du.$$

By bounding  $\Delta(u, x)$  above, we obtain that  $|\xi_{3,m_{n,x}}| \leq k_x^{1/2} \bar{\Delta}(\beta_{m_{n,x}}, x)$  and thus,  $\xi_{3,m_{n,x}} \rightarrow 0$  in probability as  $n \rightarrow \infty$  under the assumptions of Theorem 4.3. Finally,

$$\frac{\sqrt{k_x}}{\log(\beta_{m_{n,x}}/\alpha_{m_{n,x}})} \log \left( \frac{\hat{q}^{(c, \cdot)}(\alpha_{m_{n,x}}, x)}{q(\alpha_{m_{n,x}}, x)} \right) := \xi_{1,m_{n,x}} + \xi_{2,m_{n,x}} - \xi_{3,m_{n,x}}$$

converges in distribution to  $\mathcal{N} \left( \frac{1}{p_x} (\lambda(x)m(x) - \gamma_1(x)\epsilon(x)), \frac{\sigma^2(x) + \gamma_1^2(x)p_x(1-p_x)}{p_x^2} \right)$  as  $n \rightarrow \infty$ .

## References

- [1] J. Beirlant, G. Dierckx, and A. Guillou. Estimation of the extreme-value index and generalized quantile plots. *Bernoulli*, 11(6):949–970, 2005.
- [2] J. Beirlant, Y. Goegebeur, J. Teugels, and J. Segers. *Statistics of extremes*. Wiley, 2004.
- [3] J. Beirlant, A. Guillou, G. Dierckx, and A. Fils-Villetard. Estimation of the extreme value index and extreme quantiles under random censoring. *Extremes*, 10(3):151–174, 2007.

- [4] J. Beirlant, P. Vynckier, and J. L. Teugels. Tail index estimation, Pareto quantile plots, and regression diagnostics. *J. Amer. Statist. Assoc.*, 91(436):1659–1667, 1996.
- [5] B. Brahimi, D. Meraghni, and A. Necir. On the asymptotic normality of Hill’s estimator of the tail index under random censoring. *ArXiv e-prints*, February 2013.
- [6] L. de Haan and L. Peng. Comparison of tail index estimators. *Statistica Neerlandica*, 52 (1):60–70, 1998.
- [7] A. L. M. Dekkers, J. H. J. Einmahl, and L. de Haan. A moment estimator for the index of an extreme-value distribution. *Ann. Statist.*, 17(4):1833–1855, 1989.
- [8] J. H. J. Einmahl, A. Fils-Villetard, and A. Guillou. Statistics of extremes under random censoring. *Bernoulli*, 14(1):207–227, 2008.
- [9] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events*. Springer, 1997.
- [10] L. Gardes and S. Girard. A moving window approach for nonparametric estimation of the conditional tail index. *J. Multivariate Anal.*, 99(10):2368–2388, 2008.
- [11] L. Gardes and S. Girard. Conditional extremes from heavy-tailed distributions: an application to the estimation of extreme rainfall return levels. *Extremes*, 13(2):177–204, 2010.
- [12] M. I. Gomes and M. M. Neves. A note on statistics of extremes for censoring schemes on a heavy right tail. In *Luzar-Stiffler, V., Jarec, I. and Bekic, Z. (eds.), Proceedings of ITI 2010, SRCE Univ. Computing Centre Editions*, pages 539–544, 2010.
- [13] A. Lekina. Estimation non-paramétrique des quantiles extrêmes conditionnels. *Thèse de doctorat, Université de Grenoble*, 2010.
- [14] R.-D. Reiss and M. Thomas. *Statistical analysis of extreme values with applications to insurance, finance, hydrology and other fields*. Birkhäuser, 2007.
- [15] I. Weissman. Estimation of parameters and large quantiles based on the  $k$  largest observations. *J. Amer. Statist. Assoc.*, 73:812–815, 1978.

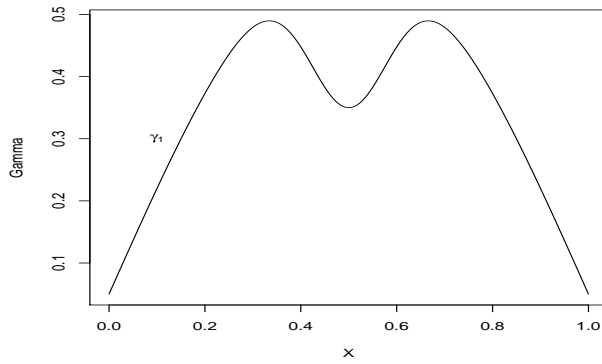
| n    | Hill estimator   |  |  | Moment estimator   |  |  | UH estimator   |  |  |
|------|--|--|--|--|--|--|--|--|--|
|      | 10%  | 25%  | 40%  | 10%  | 25%  | 40%  | 10%  | 25%  | 40%  |
| 500  | .349<br>(.037)<br>[.030]<br>[.273,.425]<br>.758 <sup>†</sup> | .350<br>(.043)<br>[.034]<br>[.270,.428]<br>.961 <sup>†</sup> | .345<br>(.045)<br>[.035]<br>[.256,.434]<br>.963 <sup>†</sup> | .323<br>(.116)<br>[.090]<br>[.081,.566]<br>.900 <sup>†</sup> | .315<br>(.134)<br>[.107]<br>[.040,.591]<br>.908 <sup>†</sup> | .317<br>(.172)<br>[.136]<br>[0*,.652]<br>.950 <sup>†</sup>   | .322<br>(.114)<br>[.090]<br>[.084,.561]<br>.945 <sup>†</sup> | .304<br>(.137)<br>[.109]<br>[.029,.580]<br>.960 <sup>†</sup> | .323<br>(.171)<br>[.135]<br>[0*,.652]<br>.964 <sup>†</sup>   |
| 1000 | .346<br>(.027)<br>[.022]<br>[.293,.400]<br>.969 <sup>†</sup> | .349<br>(.029)<br>[.023]<br>[.290,.408]<br>.986 <sup>†</sup> | .346<br>(.032)<br>[.026]<br>[.283,.409]<br>.990 <sup>†</sup> | .337<br>(.081)<br>[.065]<br>[.166,.509]<br>.959 <sup>†</sup> | .330<br>(.097)<br>[.076]<br>[.124,.535]<br>.964 <sup>†</sup> | .334<br>(.119)<br>[.093]<br>[.095,.572]<br>.970 <sup>†</sup> | .339<br>(.081)<br>[.065]<br>[.171,.508]<br>.965 <sup>†</sup> | .336<br>(.099)<br>[.079]<br>[.135,.536]<br>.968 <sup>†</sup> | .326<br>(.119)<br>[.093]<br>[.088,.561]<br>.970 <sup>†</sup> |
| 1500 | .347<br>(.021)<br>[.017]<br>[.304,.389]<br>.973 <sup>†</sup> | .348<br>(.025)<br>[.020]<br>[.300,.396]<br>.993 <sup>†</sup> | .345<br>(.030)<br>[.024]<br>[.289,.401]<br>.995 <sup>†</sup> | .344<br>(.067)<br>[.053]<br>[.208,.480]<br>.971 <sup>†</sup> | .339<br>(.071)<br>[.057]<br>[.171,.506]<br>.977 <sup>†</sup> | .335<br>(.101)<br>[.080]<br>[.122,.549]<br>.981 <sup>†</sup> | .342<br>(.067)<br>[.053]<br>[.206,.477]<br>.975 <sup>†</sup> | .340<br>(.072)<br>[.058]<br>[.183,.511]<br>.980 <sup>†</sup> | .339<br>(.101)<br>[.081]<br>[.129,.551]<br>.984 <sup>†</sup> |
| 2000 | .349<br>(.019)<br>[.015]<br>[.311,.387]<br>.987 <sup>†</sup> | .349<br>(.020)<br>[.016]<br>[.310,.389]<br>.995 <sup>†</sup> | .348<br>(.022)<br>[.017]<br>[.303,.394]<br>.998 <sup>†</sup> | .337<br>(.058)<br>[.045]<br>[.212,.463]<br>.980 <sup>†</sup> | .340<br>(.067)<br>[.052]<br>[.202,.477]<br>.984 <sup>†</sup> | .335<br>(.087)<br>[.068]<br>[.160,.509]<br>.986 <sup>†</sup> | .342<br>(.058)<br>[.046]<br>[.219,.466]<br>.981 <sup>†</sup> | .345<br>(.068)<br>[.053]<br>[.209,.480]<br>.984 <sup>†</sup> | .337<br>(.087)<br>[.068]<br>[.164,.512]<br>.987 <sup>†</sup> |

**Table 1 (simulation results for  $\gamma_1(x)$ ).** For each configuration of the simulation parameters ( $n$ ,  $c$ , tail index estimator), the first line gives the averaged value of the  $R = 1000$  estimates of  $\gamma_1(x)$ . (·): empirical root MSE of the estimates. [·]: empirical mean absolute error. [·, ·]: 95%-level asymptotic confidence interval for  $\gamma_1(x)$  (a \* indicates that the lower bound was negative and thus truncated to 0). †: empirical coverage probability.

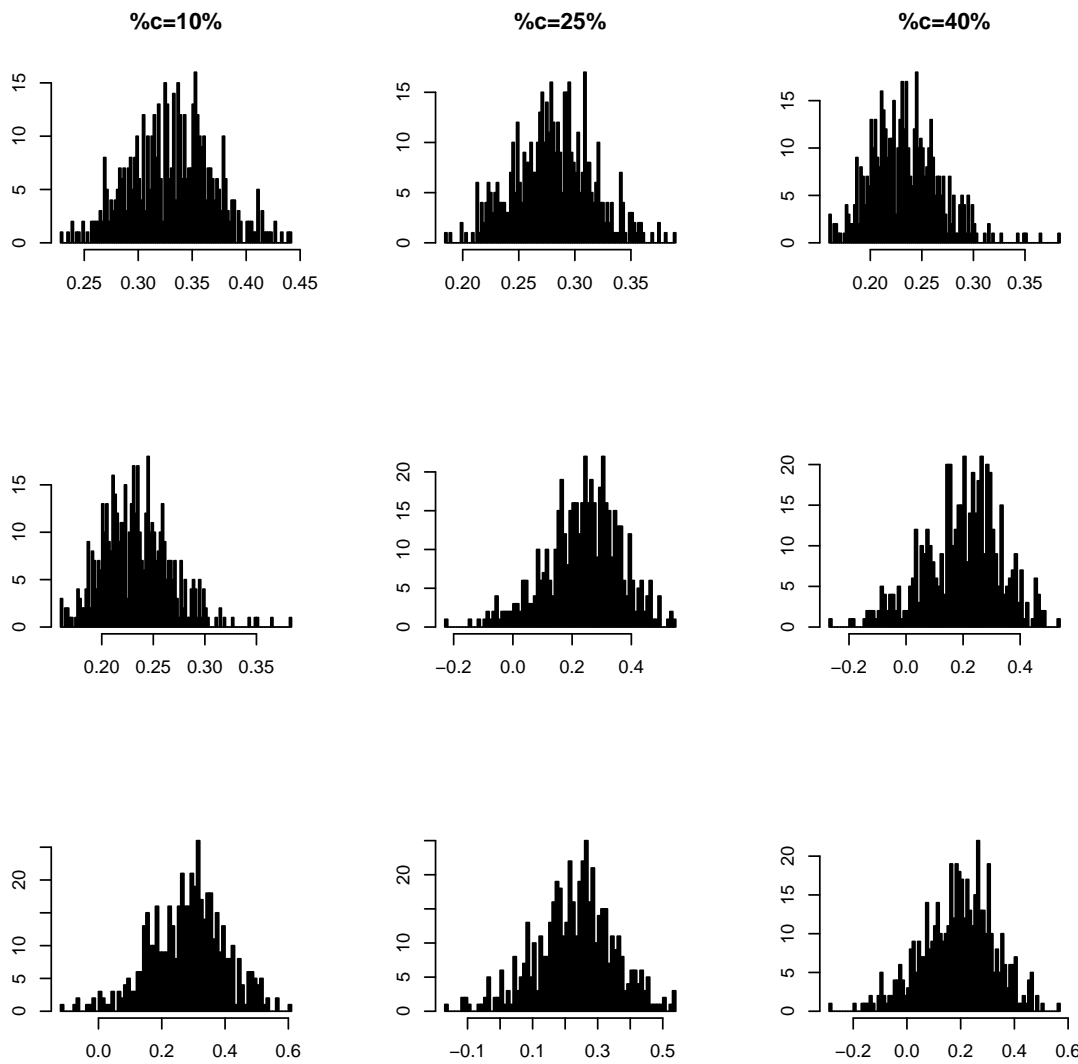
| n    | Hill estimator   |  |  | Moment estimator  |   |  | UH estimator   |   |  |
|------|--|--|--|---|---|--|--|---|--|
|      | 10%  | 25%  | 40%  | 10%   | 25%   | 40%  | 10%  | 25%   | 40%  |
| 500  | 19.777<br>(.258)<br>[.326]<br>[16.00,25.88]<br>0.594†  | 20.225<br>(.265)<br>[.333]<br>[15.90,27.77]<br>0.936†  | 20.072<br>(.310)<br>[.383]<br>[15.56,28.25]<br>0.970†  | 27.490<br>(.807)<br>[1.008]<br>[15.81,105.03]<br>0.680† | 27.343<br>(.877)<br>[1.122]<br>[14.76,184.95]<br>0.880† | 27.279<br>(1.164)<br>[1.431]<br>[14.25,317.34]<br>0.882† | 25.313<br>(.806)<br>[1.007]<br>[14.56,96.71]<br>0.230† | 29.046<br>(.892)<br>[1.154]<br>[15.82,176.50]<br>0.921† | 37.537<br>(1.213)<br>[1.508]<br>[20.05,292.11]<br>0.926† |
| 1000 | 19.381<br>(.182)<br>[.226]<br>[16.54,23.39]<br>0.708†  | 19.960<br>(.206)<br>[.259]<br>[16.71,24.77]<br>0.971†  | 20.086<br>(.222)<br>[.280]<br>[16.55,25.53]<br>0.989†  | 24.065<br>(.565)<br>[.712]<br>[15.60,52.60]<br>0.863†   | 22.762<br>(.671)<br>[.843]<br>[14.00,60.69]<br>0.942†   | 27.813<br>(.820)<br>[1.022]<br>[16.81,80.31]<br>0.948†   | 23.189<br>(.562)<br>[.709]<br>[15.03,50.68]<br>0.262†  | 23.465<br>(.676)<br>[.841]<br>[14.53,60.91]<br>0.958†   | 27.695<br>(.826)<br>[1.025]<br>[16.82,78.30]<br>0.961†   |
| 1500 | 19.841<br>(.142)<br>[0.177]<br>[17.75,22.47]<br>0.910† | 19.981<br>(.161)<br>[0.199]<br>[17.21,23.59]<br>0.990† | 19.905<br>(.179)<br>[0.223]<br>[17.26,23.70]<br>0.992† | 22.506<br>(.468)<br>[0.597]<br>[16.32,36.19]<br>0.904†  | 21.621<br>(.534)<br>[0.666]<br>[14.39,43.37]<br>0.957†  | 22.762<br>(.639)<br>[0.818]<br>[15.15,45.72]<br>0.974†   | 22.584<br>(.473)<br>[0.600]<br>[16.38,36.32]<br>0.972† | 23.106<br>(.544)<br>[0.674]<br>[15.49,45.42]<br>0.978†  | 26.684<br>(.638)<br>[0.684]<br>[17.97,51.72]<br>0.981†   |
| 2000 | 19.887<br>(.131)<br>[.164]<br>[17.79,21.53]<br>0.922†  | 19.841<br>(.142)<br>[.177]<br>[17.42,23.04]<br>0.992†  | 20.048<br>(.160)<br>[.202]<br>[17.60,23.28]<br>0.994†  | 21.889<br>(.400)<br>[.503]<br>[15.85,35.36]<br>0.953†   | 22.506<br>(.468)<br>[.597]<br>[15.61,40.28]<br>0.963†   | 22.132<br>(.616)<br>[.766]<br>[15.28,40.07]<br>0.978†    | 21.159<br>(.397)<br>[.500]<br>[15.32,34.18]<br>0.277†  | 22.584<br>(.473)<br>[.600]<br>[15.71,40.13]<br>0.981†   | 24.696<br>(.624)<br>[.771]<br>[17.20,43.72]<br>0.983†    |

**Table 2 (simulation results for  $q(1/5000, .5)$ ).** For each configuration of the simulation parameters ( $n$ ,  $c$ , tail index estimator), the first line gives the averaged value of the  $R = 1000$  estimates of  $q(1/5000, .5)$ .  $(\cdot)$ : empirical root MSE.  $[\cdot]$ : empirical mean absolute error.  $[\cdot, \cdot]$ : 95%-level asymptotic confidence interval for  $q(1/5000, .5)$ .  $\dagger$ : empirical coverage probability.

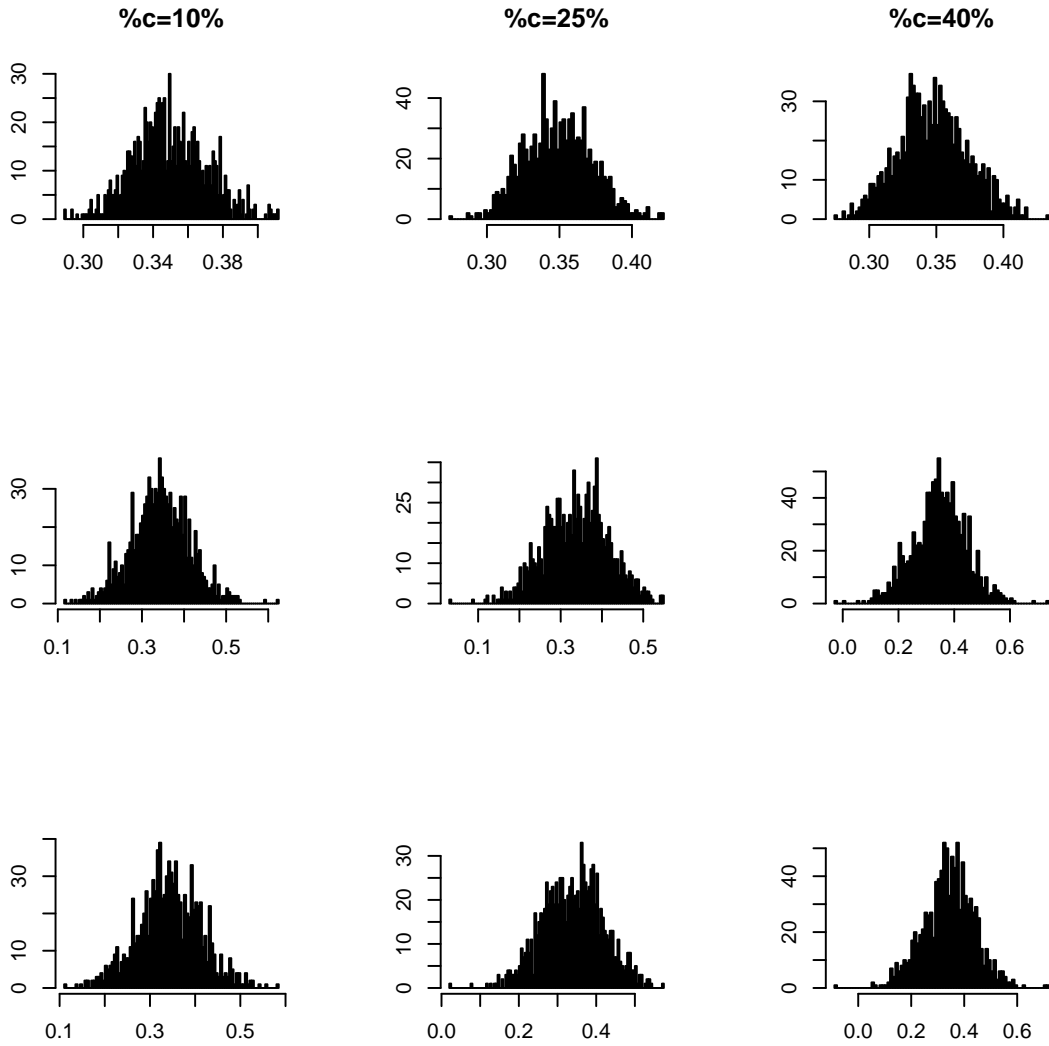




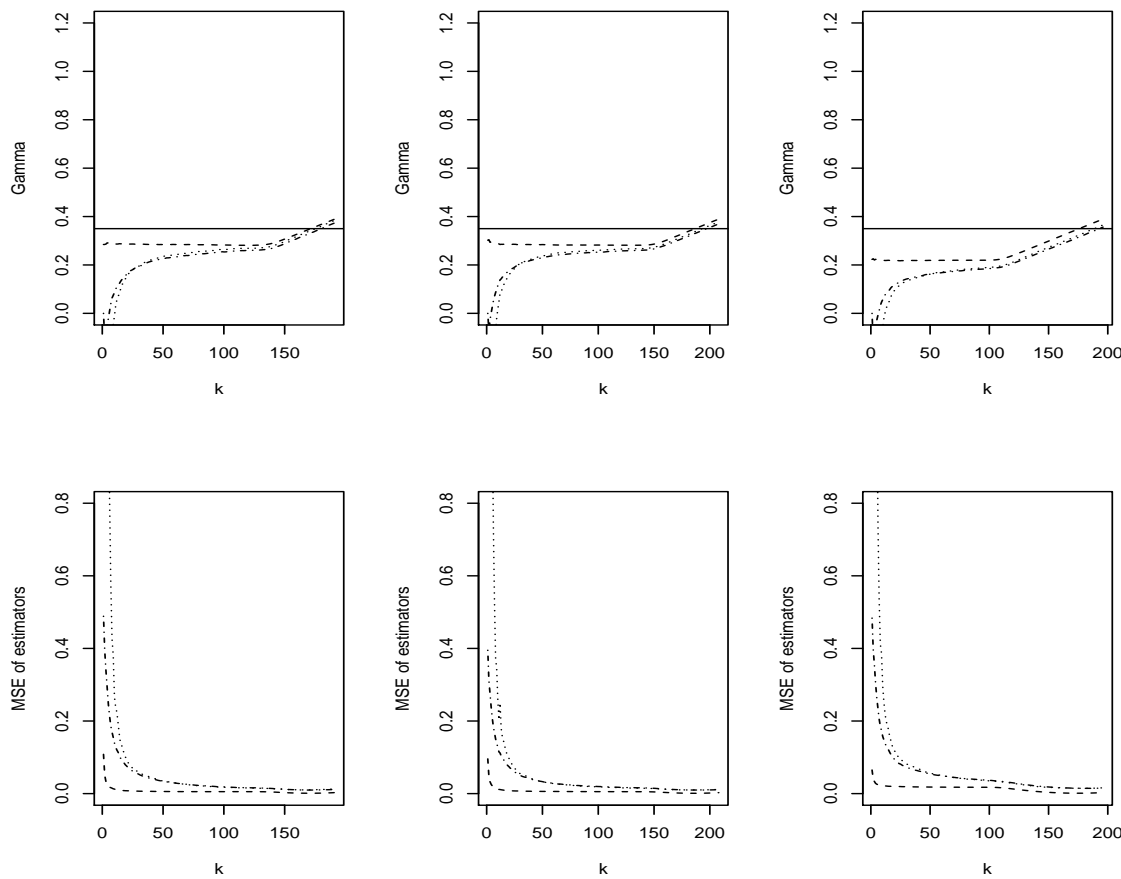
**Figure 1.** Pattern of the function  $\gamma_1(\cdot)$  on  $[0, 1]$ .



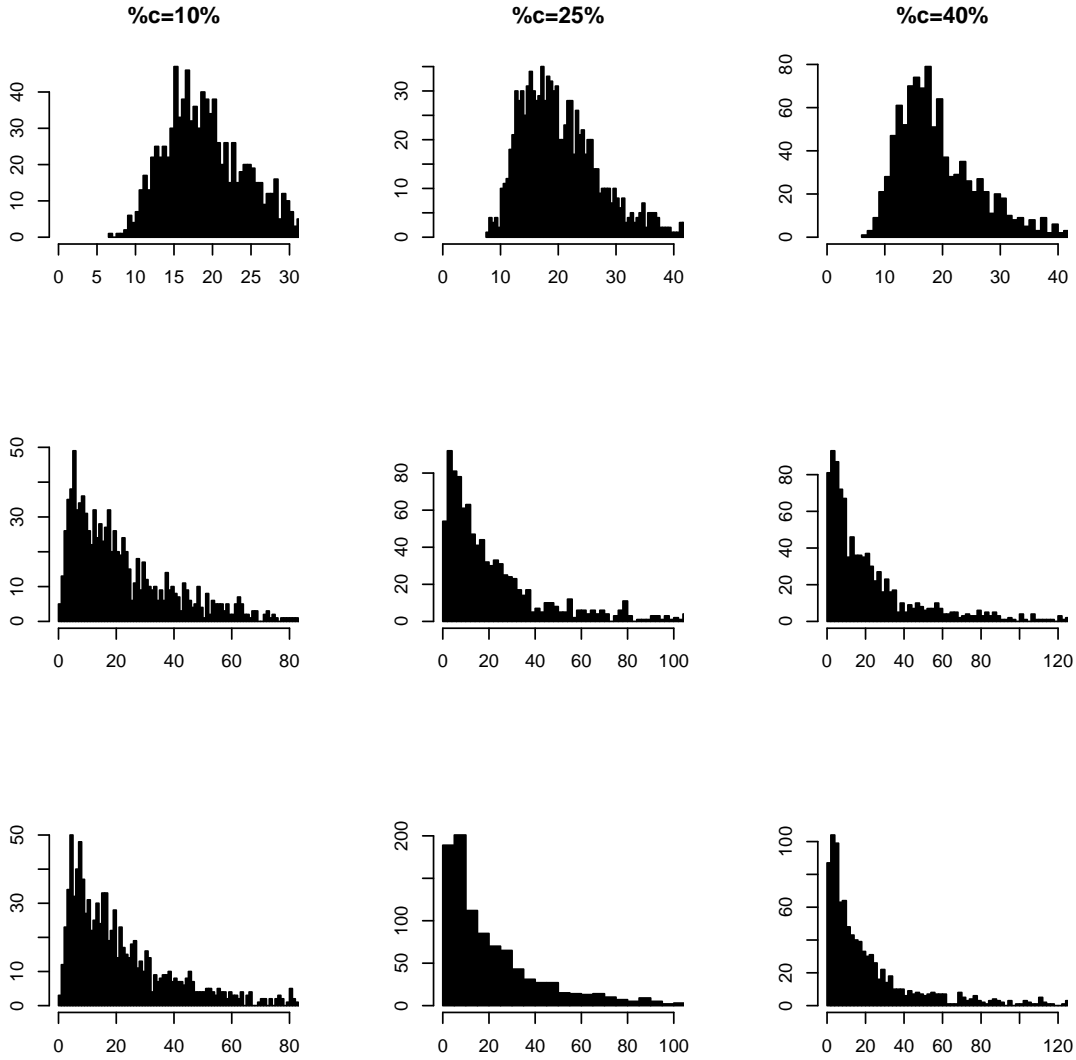
**Figure 2.** Histograms of the  $R = 1000$  Hill (1st line), moment (2nd line), and UH (3rd line) estimates of the tail index at  $x = .5$  ( $\gamma_1(.5) = .35$ ), for  $c = 10\%$  (left column),  $c = 25\%$  (middle),  $c = 40\%$  (right). The sample size is 500.



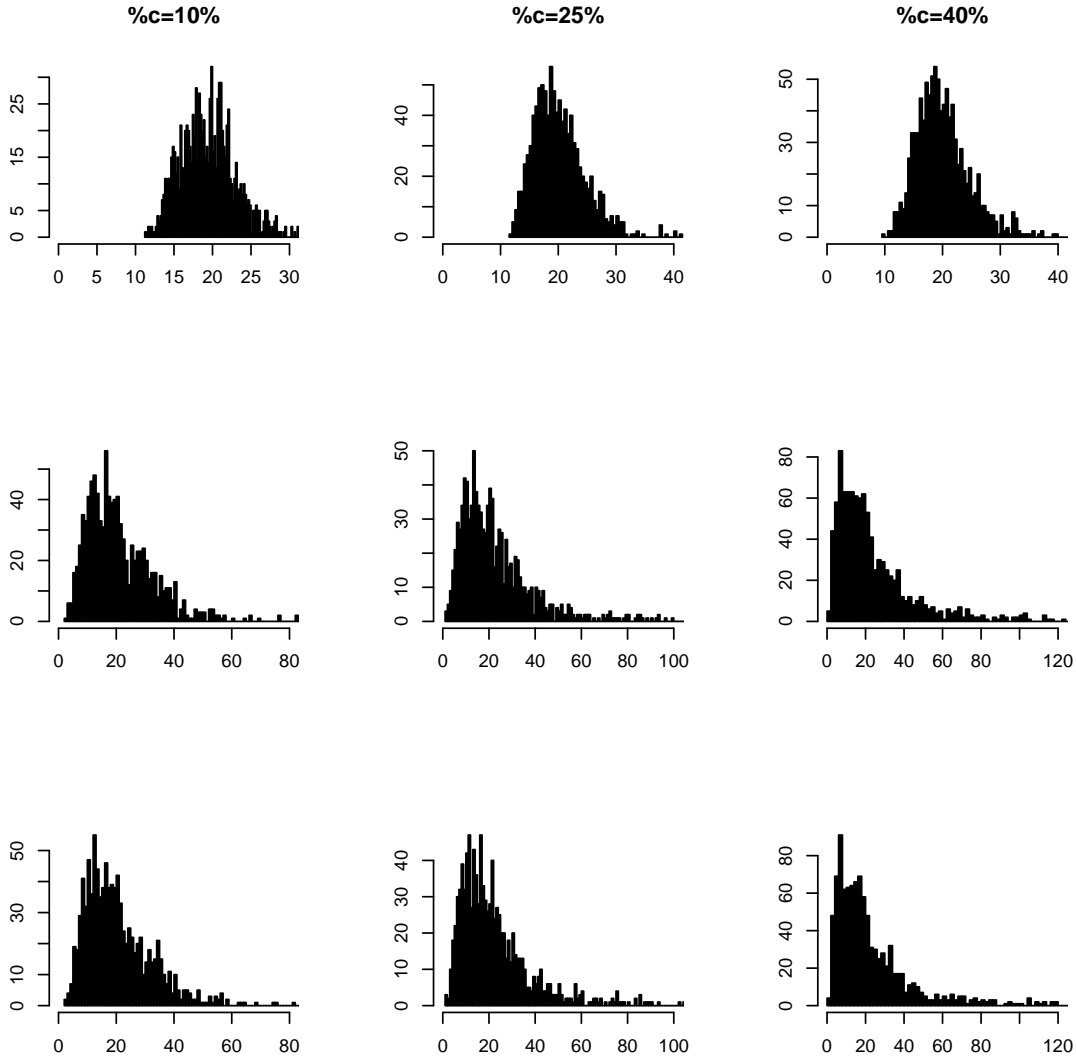
**Figure 3.** Histograms of the  $R = 1000$  Hill (1st line), moment (2nd line), and UH (3rd line) estimates of the tail index at  $x = .5$  ( $\gamma_1(.5) = .35$ ), for  $c = 10\%$  (left column),  $c = 25\%$  (middle),  $c = 40\%$  (right). The sample size is 1500.



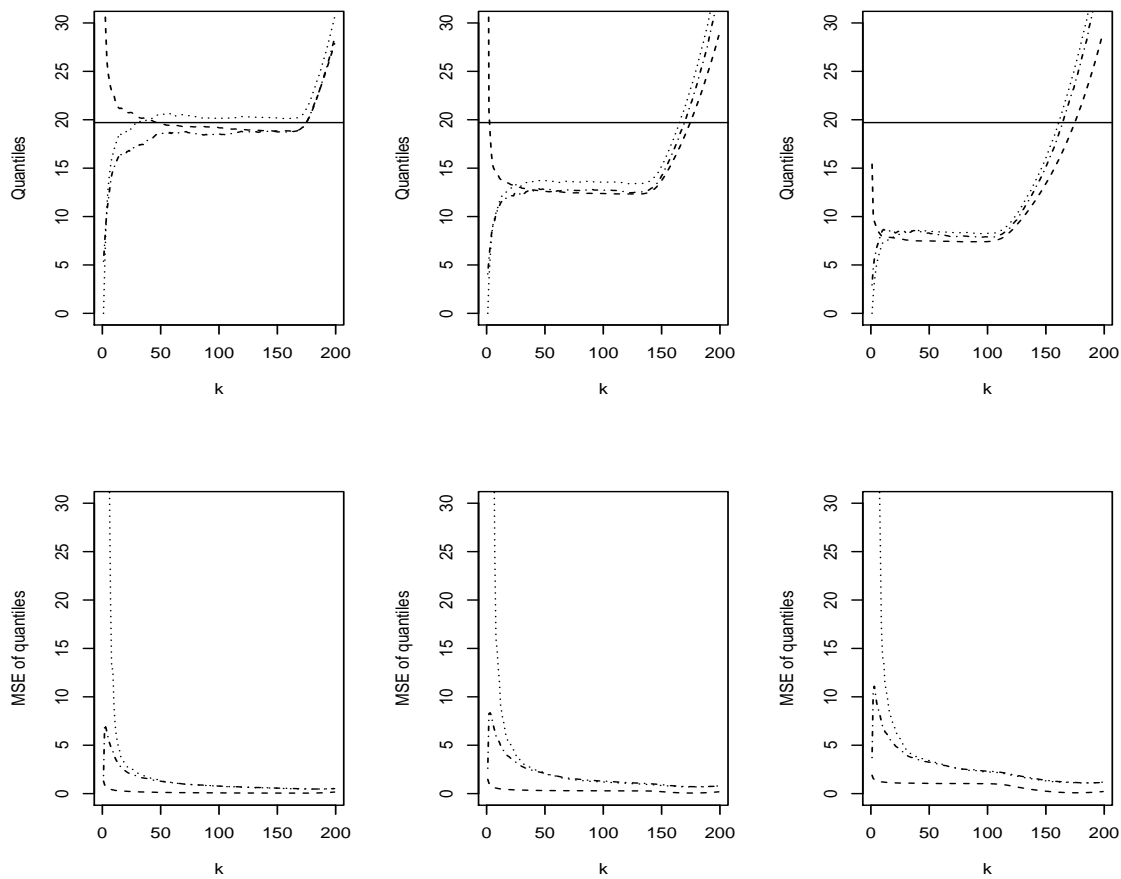
**Figure 4.** Averaged value (upper-panel) and empirical mean square error (lower panel) of the  $R$  estimates of  $\gamma_1(.5) = .35$ , for the Hill (dashed line), moment (dotted line), and UH (dash-dotted line) estimators, for  $c = 10\%$  (left column),  $c = 25\%$  (middle),  $c = 40\%$  (right).  $n = 1000$ . The true value  $\gamma_1(.5) = .35$  is represented as the black constant line (upper panel).



**Figure 5.** Histograms of the  $R = 1000$  estimates of  $q(1/5000, .5) \approx 19.70786$ , based on the Hill (1st line), moment (2nd line), and UH (3rd line) estimates of the conditional tail index, for  $c = 10\%$  (left column),  $c = 25\%$  (middle),  $c = 40\%$  (right). The sample size is 500.



**Figure 6.** Histograms of the  $R = 1000$  estimates of  $q(1/5000, .5) \approx 19.70786$ , based on the Hill (1st line), moment (2nd line), and UH (3rd line) estimates of the conditional tail index, for  $c = 10\%$  (left column),  $c = 25\%$  (middle),  $c = 40\%$  (right). The sample size is 1500.



**Figure 7.** Averaged value (upper-panel) and empirical mean square error (lower panel) of the  $R$  estimates of  $q(1/5000, .5) \approx 19.70786$ , based on the Hill (dashed line), moment (dotted line), and UH (dash-dotted line) estimators of the tail index, for  $c = 10\%$  (left column),  $c = 25\%$  (middle),  $c = 40\%$  (right).  $n = 1000$ . The true  $q(1/5000, .5)$  is represented as the black constant line (upper panel).